

- Note:**
1. Question 1 is compulsory
 2. Answer any 4 out of the remaining questions.
 3. Answers to sub questions must be written together

Q.1 (a) What are differences between Data Warehouse and Data Mart ? (05)

(b) For a Supermarket Chain consider the following dimensions, namely Product, store, time, promotion. The schema contains a central fact table, sales facts with three measures unit_sales, dollars_sales and dollar_cost. Design star schema for this application. (05)

(c) Calculate the maximum number of base fact table records for warehouse with the following values given below : (05)

- Time period: 5 years
- Store: 300 stores reporting daily sales
- Product: 40,000 products in each store (about 4000 sell in each store daily)

(d) Illustrate how the supermarket can use clustering methods to improve sales. (05)

Q2. Define the following terms by giving examples

- (a) Factless fact tables
- (b) Snowflake Schema
- (c) Web Structure Mining
- (d) Concept Hierarchy

(5 X 4 =20)

Q.3 (a) Apply Agglomerative Hierarchical Clustering and draw single link and average link dendrogram for the following distance matrix. (10)

	A	B	C	D	E
A	0	2	6	10	9
B	2	0	3	9	8
C	6	3	0	7	5
D	10	9	7	0	4
E	9	8	5	4	0

(b) Explain the Page Rank technique with algorithm. (10)

Q 4.(a) Consider a data warehouse for a hospital, where there are three dimensions:

(1) Doctor (2) Patient (3) Time; and two measures: (1) Count & (2) Fees;

For this example create a OLAP cube and describe the following OLAP operations:

(1) Slice (2) Dice (3) Rollup (4) Drill Down (5) Pivot (10)

(b) Consider the following transaction database:

TID	Items
01	A, B, C, D
02	A, B, C, D, E, G
03	A, C, G, H, K
04	B, C, D, E, K
05	D, E, F, H, L
06	A, B, C, D, L
07	B, I, E, K, L
08	A, B, D, E, K
09	A, E, F, H, L
10	B, C, D, F

Apply the Apriori algorithm with minimum support of 30% and minimum confidence of 70%, and find all the association rules in the data set. (10)

Q 5.(a) A simple example from the stock market involving only discrete ranges has Profit as categorical attribute, with values {up, down}. and the training data is:

AGE	COMPETITION	TYPE	PROFIT
Old	Yes	software	Down
Old	No	software	Down
Old	No	hardware	Down
Mid	Yes	software	Down
Mid	Yes	hardware	Down
Mid	No	hardware	Up
Mid	No	software	Up
New	Yes	software	Up
New	No	hardware	Up
New	No	software	Up

Apply the decision tree algorithm and show the generated rules. (10)

(b) Describe the steps of the ETL (Extract - Transform - Load) cycle. (10)

Q6. (a) Define multidimensional and multilevel association mining. (10)

(b) Explain the role of Meta data in a data warehouse. (10)

Q7. Write detailed notes on;

(a) Data Warehouse Architecture

(b) K-Means Clustering

(10 X 2 = 20)